We Need To Talk About Metabarcoding

Identification of *Phytophthora* species in environmental samples

Leighton Pritchard[1], Peter Cock[1], Peter Thorpe[1,2], Beatrix Clark[1], Eva Randall[1], Sarah Green[3], David Cooke[1]

[1]The James Hutton Institute, Invergowrie, Dundee, DD2 5DA
[2]School of Medicine, University of St Andrews, North Haugh, St Andrews, UK
[3]Forest Research, Bush Estate, Roslin, Edinburgh, UK

Email: leighton.pritchard@hutton.ac.uk

Forest Research

Download/cite this poster
DOI:10.6084/m9.figshare.8832296.v1

The James Hutton Institute

Introduction

The PHYTO-THREATS project aims to address risks to UK forest and woodland ecosystems from pathogenic *Phytophthora* species, which impact forests and woodlands on a global scale. Trade in ornamental plants, which may harbour invasive pathogens, is a driver of disease outbreaks, so we are examining the distribution and diversity of *Phytophthora* in UK plant nursery systems, to identify good practices that restrict pathogen spread.

ITS1 metabarcoding is a modern, high-throughput, more sensitive alternative to conventional culturing and baiting, for identification of *Phytophthora* in environmental samples. This approach can potentially detect all species (known or unknown) of the target genus that are present in the sample. However, the method's extreme sensitivity and prevalence of sequencing artefacts presents problems for identification and interpretation, especially in regulatory contexts.

We present evidence from a study that demonstrates the rigorous application of ITS1 metabarcoding using Illumina sequencing for detection and identification of *Phytophthora* species in environmental samples.

We also present THAPBI-pict, a new software tool for *Phytophthora* ITS1 metabarcoding sequence classification.

THAPBI-pict website

PHYTO-THREATS website

Acknowledgements
This work is supported by a grant funded jointly by the Biotechnology and Biological Sciences Research Council, the Department for Environment, Food and Rural affairs, the Economic and Social Research Council, the Forestry Commission, the Natural Environment Research Council and the Scottish Government, under the Tree Health and Plant Biosecurity Initiative.

We are grateful to Jenny Morris and Peter Hedley for their assistance with sequencing.

Scottish Government
Riaghaltas na h-Alba
gov.scot

Methods

A consensus ITS1 sequence was obtained from an alignment of *Phytophthora* ITS1 reference sequences. This was shuffled, preserving single-base and dyad composition, to generate four synthetic ITS1 variant sequences. Synthetic DNA of each control sequence was purchased from Integrated DNA Technologies, and extended to include primer-binding sites for the standard nested PCR protocol with 18Ph2F/5.8S-1R and ITS6/5.8S-1R primers.

The resulting sequences were mixed as six pools, with alternative 1:10:100:1000 dilutions of the four synthetic ITS1 variants, to represent different biological communities. The pools were diluted at 1X, 10X and 100X to represent a range of input biomasses for the same community.

The synthetic sample pools were included on a plate of environmental samples and single-isolate controls for standard ITS1 metabarcoding using a nested PCR protocol with a proof-reading enzyme. All samples were barcoded with dual indices and Illumina sequencing adapters, and pooled for 2x250bp paired-end sequencing on a single Illumina MiSeq flow cell, at the James Hutton Institute.

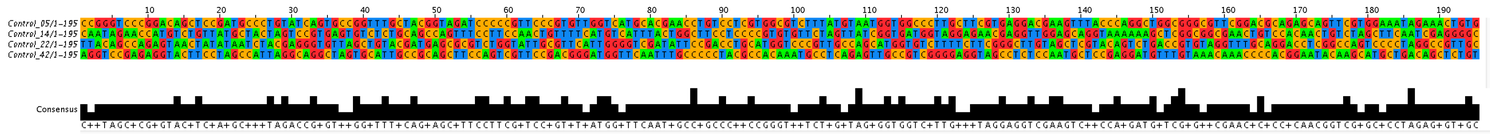

Figure 1 Synthetic control ITS1 sequences

Results

The synthetic control sequences have the same base and dyad composition as a consensus ITS1 sequence, therefore we assume that they behave similarly to ITS1 sequences sampled from the environment. We expect that: observed amplified sequence variation for synthetic controls represents artefactual sequence variation for environmental ITS1 sequences; (relative) quantification of synthetic sequences is representative of quantification of environmental ITS1 sequences; environmental *Phytophthora* ITS1 sequence present in a synthetic control sample represents cross-contamination of a typical sample during amplification, barcoding, and sequencing.

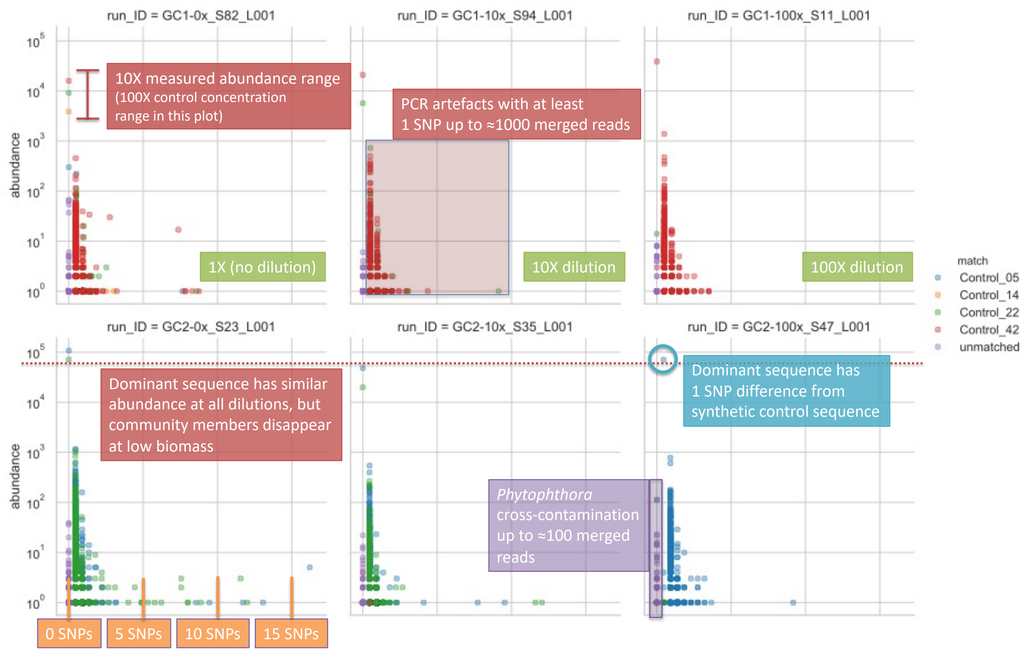

Figure 2 Merged read abundance vs SNP distance from control sequence for two pools at 1X, 10X and 100X dilution.

We find:

1. **All input sequences can generate large numbers of sequence variants in PCR**
2. **The dominant sequence variant has similar absolute abundance independent of initial biomass**
3. **Abundances are not absolutely or relatively quantitative**
4. **With low biomass, the most abundant sequence may be an artefact**
5. **Cross-contamination produces sequences that occur at 100-1000 abundance**.

We have produced a new software tool for *Phytophthora* ITS1 metabarcoding analysis, called THAPBI-pict. This tool also provides a manually-curated companion reference database containing ITS1 sequences from single-isolate *Phytophthora* controls taken through the ITS1 sequencing process. THAPBI-pict uses synthetic controls – where available – to dynamically threshold calls of species presence/absence on a plate-by-plate basis, for increased accuracy.

This tool is under active development and can be obtained at https://github.com/peterjc/thapbi-pict/

Conclusions

- **Our results have general implications for metabarcoding beyond identification of *Phytophthora* by ITS1.**
- With appropriate controls and interpretation, ITS1 metabarcoding is a useful tool for identification of species that are present in a sample, but it is not definitive for determining absence of a species, or for quantitation.
- There is evidence for extensive artefact generation by PCR, and cross-contamination between samples.
- Low-abundance sequences (≤100-1000 merged reads) should be treated with caution.
- We consider that, to guard against over-interpretation of data, adequate negative and synthetic controls should be run on a plate-by-plate basis to account for PCR artefacts and cross-contamination.